Carnegie Mellon University

HERSHEY
THE HERSHEY COMPANY

Hershey (4)

PRODUCT SPACE OPTIMIZATION TO BOOST SALES

Enterprise Data Science

Agenda

- Team introduction
- Hershey current statistics
- Hershey products
- Project goals
- Resources
- Solution
- Challenges
- Business Impact
- Next Steps/ Future Work

Team Introduction

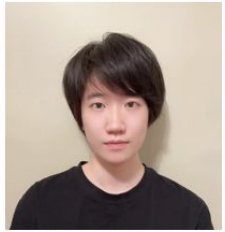
Urvish Thakker
Project Manager

Rohit Govindan
Financial Manager

Yucheng Li
Chief Systems Administrator

Noopur Latkar
Process organizer

Yashvi Thakkar
Data Scientist



Jinghuiyu Yang
Data Scientist

Hershey Team

- **Sabrina Li**
- **Ian Smith**

Faculty Advisor

Prof Dr. Murli Viswanathan

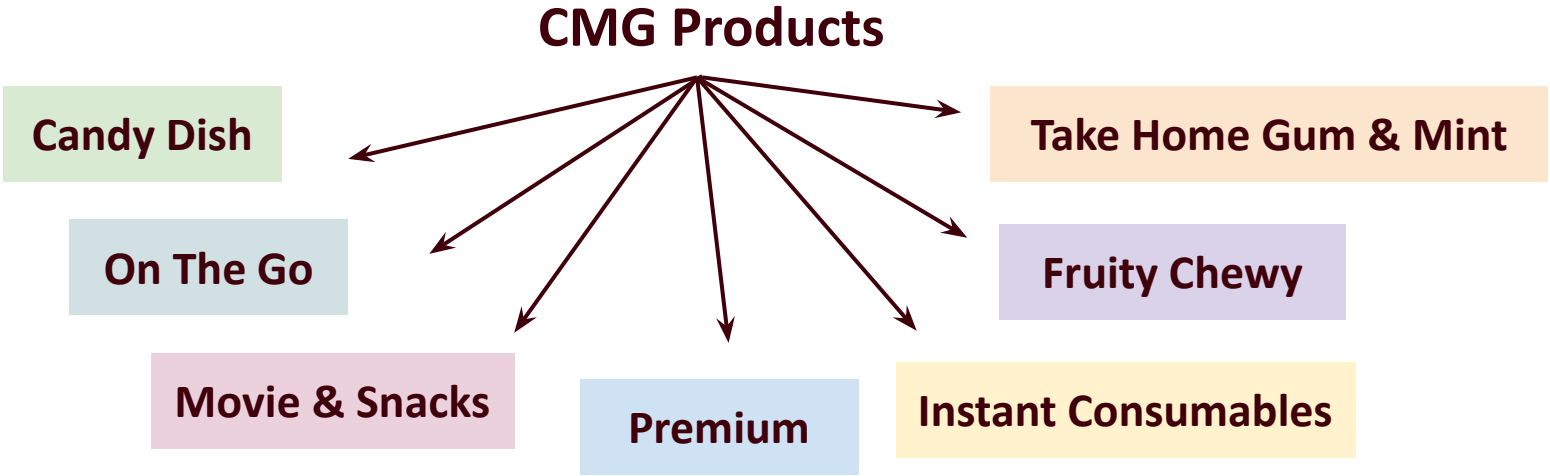

~$9B
NET SALES
~18,000
EMPLOYEES AROUND THE WORLD
PRODUCTS AVAILABLE IN
OVER 70
COUNTRIES AROUND THE WORLD
PRODUCTS MANUFACTURED
IN 7
COUNTRIES

Main Products

HERSHEY
THE HERSHEY COMPANY

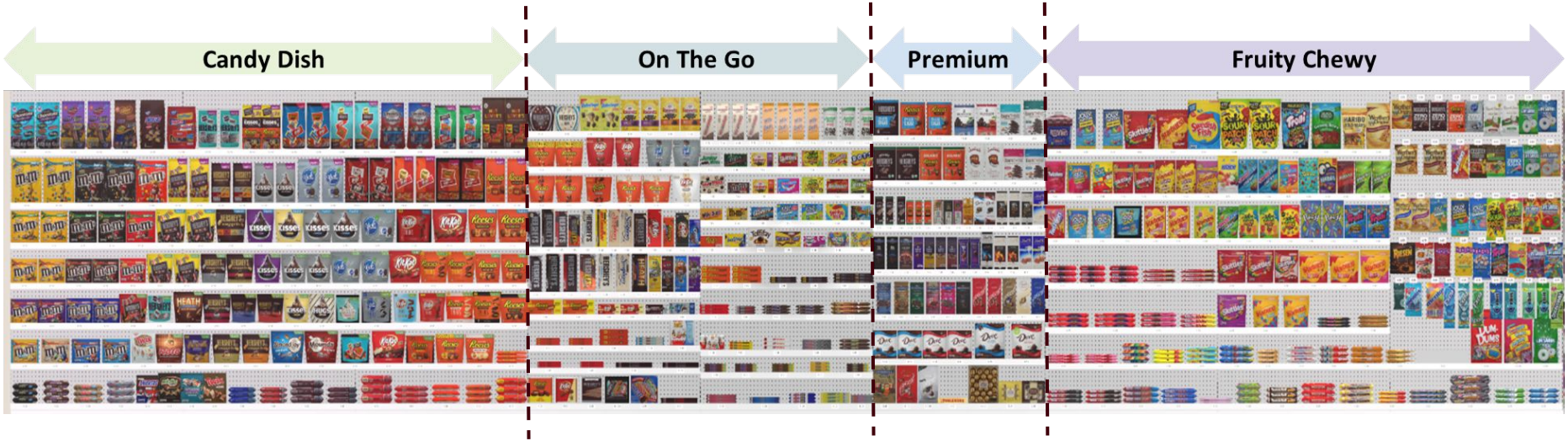
CMG : Candy Mint Gum

Usage Occasions



Distinct groups based off when consumers are **most likely to purchase** each individual product

Focusing on **4 main usage occasions** present in all main candy aisles :

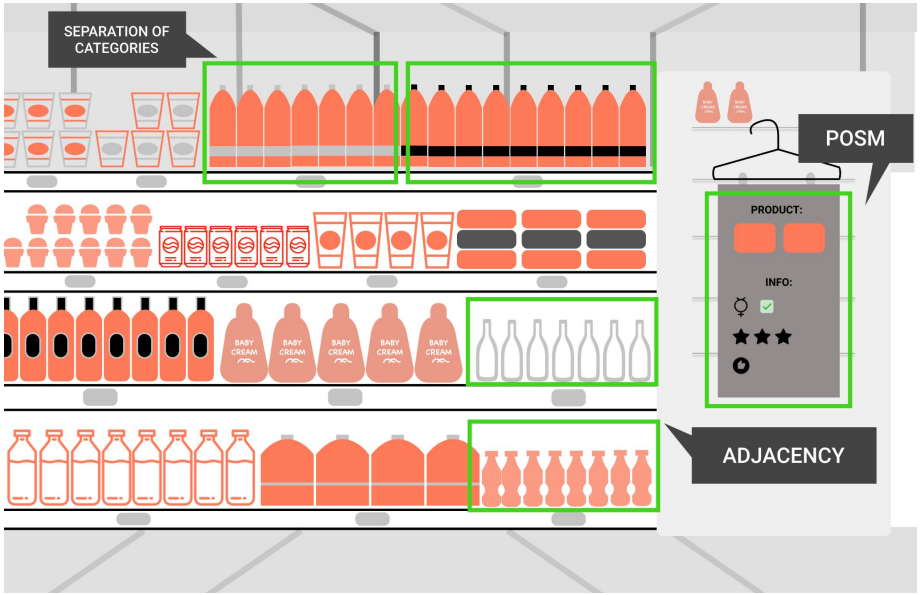


→ not included

Project Goals: Space Optimization

HERSHEY
THE HERSHEY COMPANY

Determine the **optimal amount** of space for each **usage occasion** in terms of percentage.

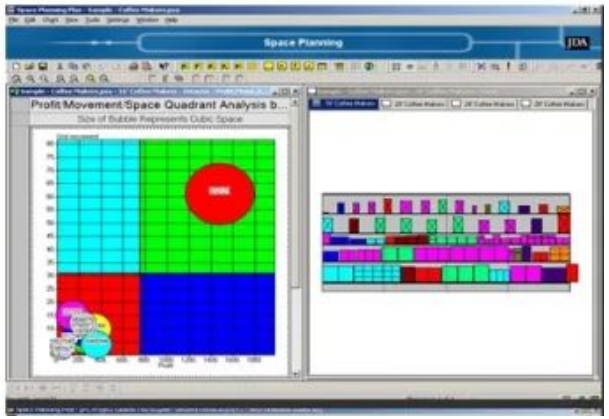


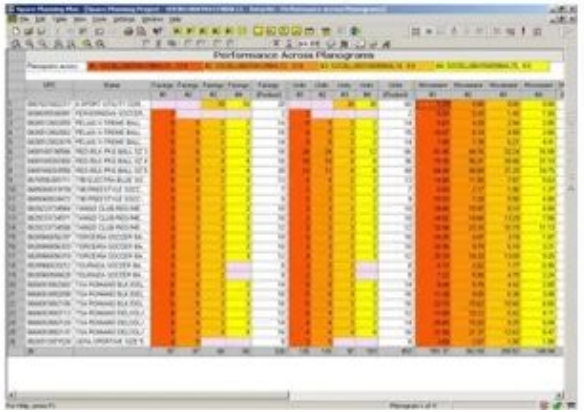
Subsidiary Project Goals

- Prescriptive Analytics
- Generalizability
- Metrics Tracking

Resources: Data

- Store Level Sales Data for **2 retailers** spanning over **4 timeframes**
- Store level Planogram Data for both retailers for 4 timeframes
- Product Attributes Data
- Access to present codebase/model employed currently for space optimization





Exploratory Data Analysis

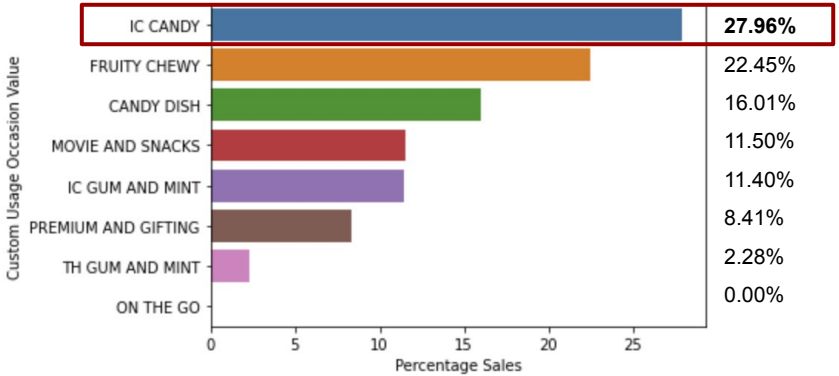
Top Selling Hershey products

	Product	Dollar Sales	UPC 10 digit
0	REESES PBC 2.8 OZ KNG MCHOC PNTBTR CUP	576055323.073	17601112125
1	HSY MILK CHOC 9.3 OZ FULLSZ MCHOC BAR	410273668.219	6800058040
2	M&M PEANUT 3.27 OZ KNG MCHOC PNT PCE	373437959.741	16000051534
3	SNICKERS 1.86 OZ STD MCHOC PNT CRML NGT BAR	348460108.827	16000206230
4	SNICKERS 3.29 OZ KNG MCHOC PNT CRML NGT BAR	346660962.123	4000000263
5	KINDER JOY .7 OZ MCHOC CRM CRSP WFR PCE	343254762.852	7920250508
6	TWIX 3.02 OZ KNG MCHOC CRML CKY BAR	289384187.625	12000054728
7	REESES PBC 1.5 OZ STD MCHOC PNTBTR CUP	281928151.483	10200132250
8	M&M PEANUT 1.74 OZ STD MCHOC PNT PCE	239121821.613	20000160250
9	M&M PEANUT 10.7 OZ MCHOC PNT PCE	232781112.899	4000051305

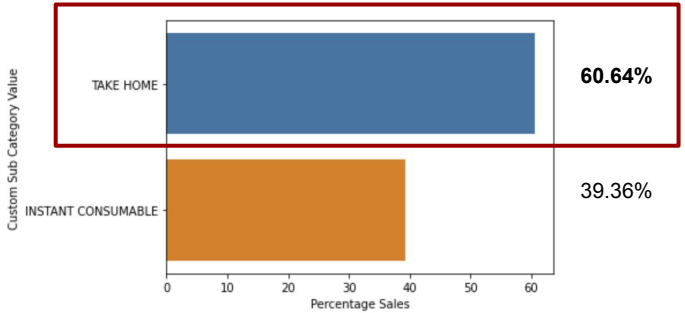
1. **Reeses Peanut Butter Cup** **$576 M**
2. Hershey Milk Choc Bar $410 M
3. M&M Peanut $373 M

Exploratory Data Analysis

Top Selling Categories (Usage Occasions)



Top Selling Sub-categories



Exploratory Data Analysis

Analyzing **Sales** by **Usage Occasion**

- Highest **selling products** per category

1. Candy Dish

Sales

$0.23 B

$0.15 B

2. Movie and Snacks

Sales

$0.40 B

$87 M

Exploratory Data Analysis

Analyzing **Sales** by **Usage Occasion**

- Highest **selling products** in usage Occasion

3. Premium Gifting

Sales

$0.10 B

$77 M

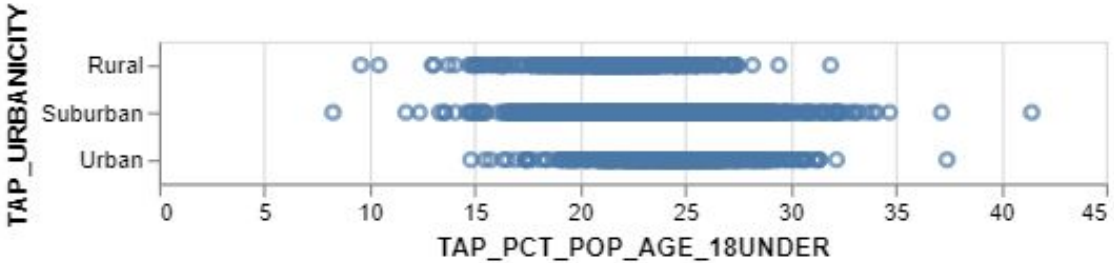
4. Fruity Chewy

Sales

$0.14 B

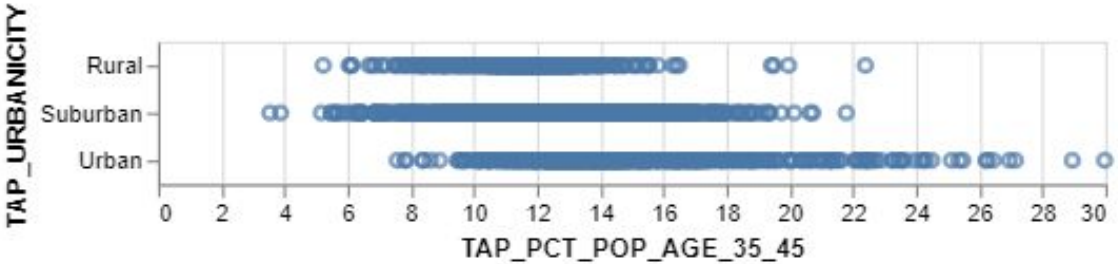
$0.12 B

Exploratory Data Analysis - Retailer A

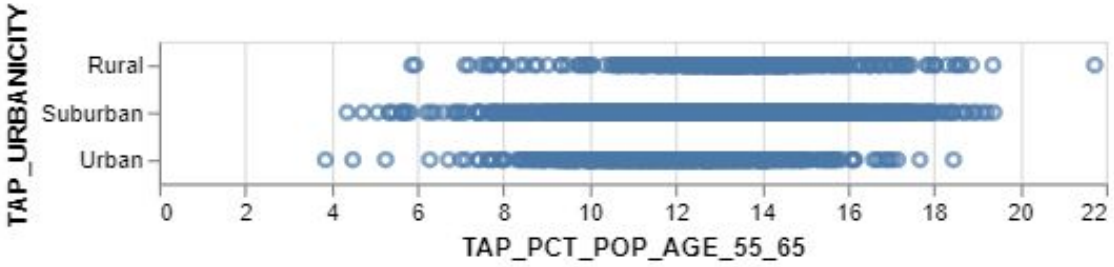


Suburban Areas, highest percentage of children

Suburban Areas, greater density of adults (35 - 45 years)

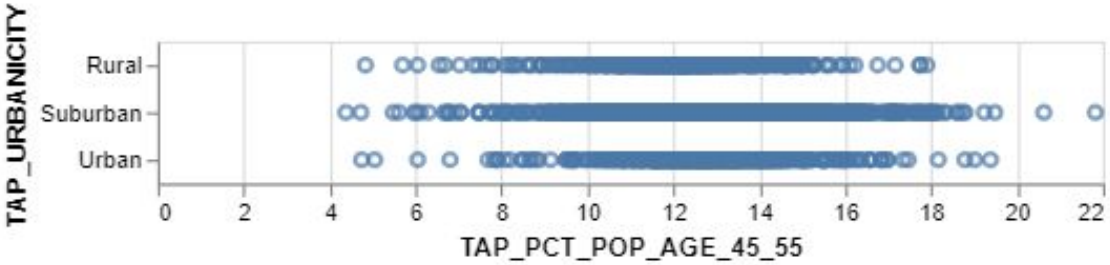


Exploratory Data Analysis - Retailer A

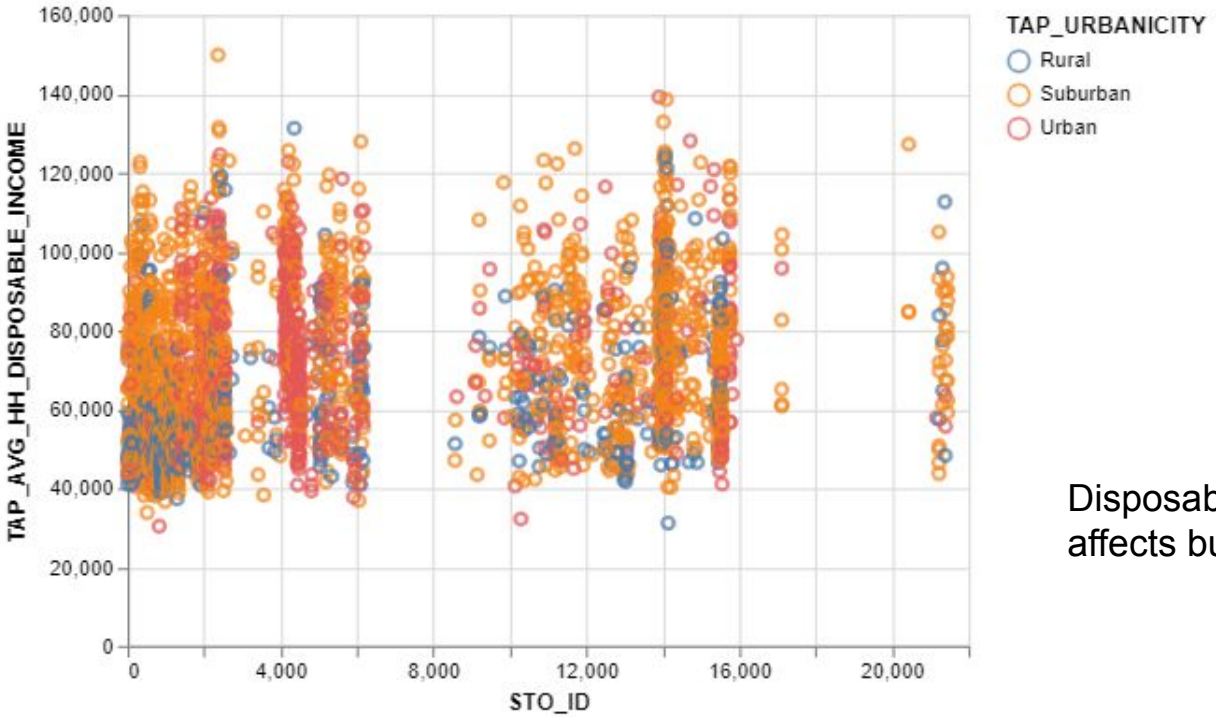


Suburban Areas, concentrated with middlehood adults

Probable parents, and adults contributing sales of candies and premium chocolates.

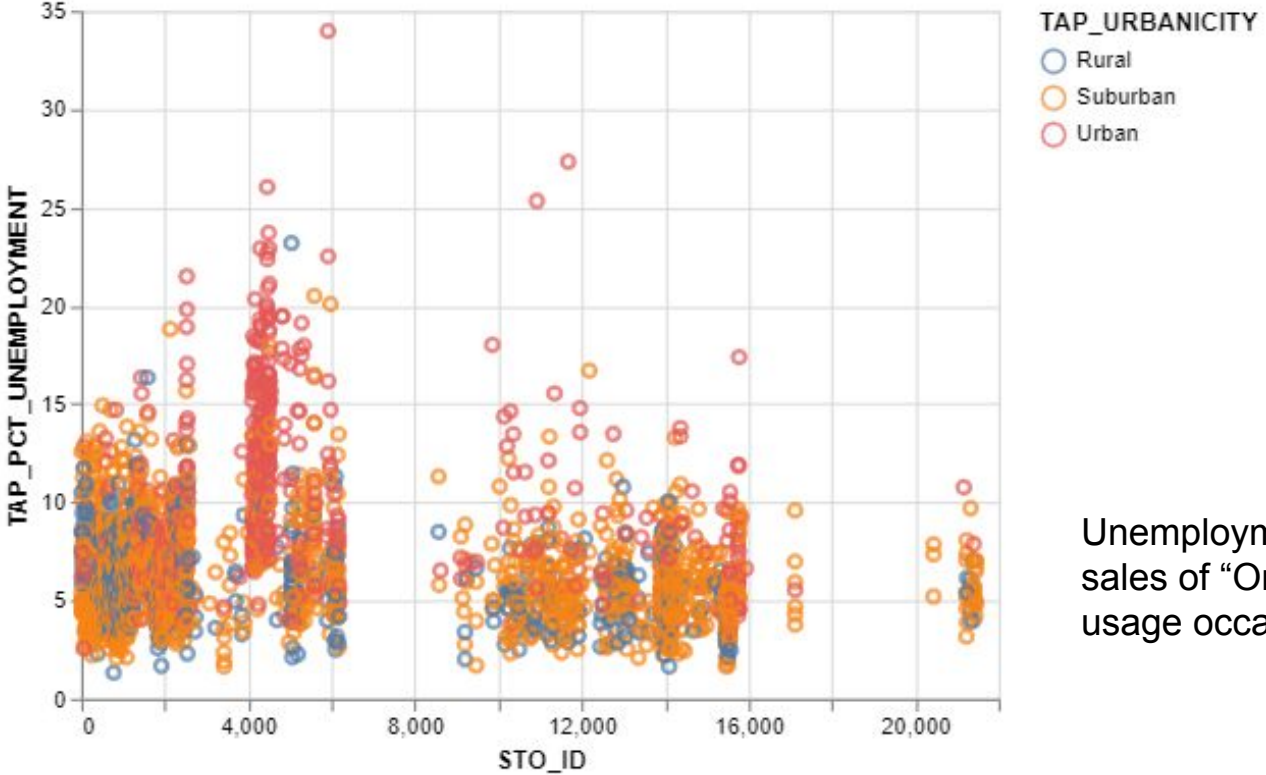


Exploratory Data Analysis - Retailer A



Disposable Income affects buying patterns

Exploratory Data Analysis - Retailer A



Unemployment drives sales of “On the Go” usage occasion

Exploratory Data Analysis - Retailer A

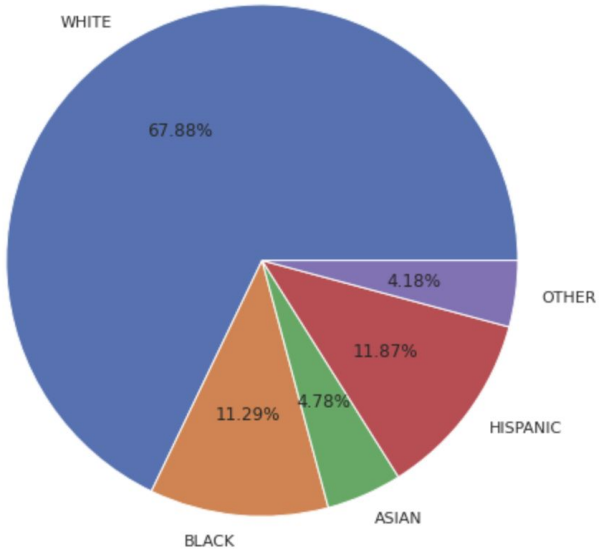
```
df.columns
```

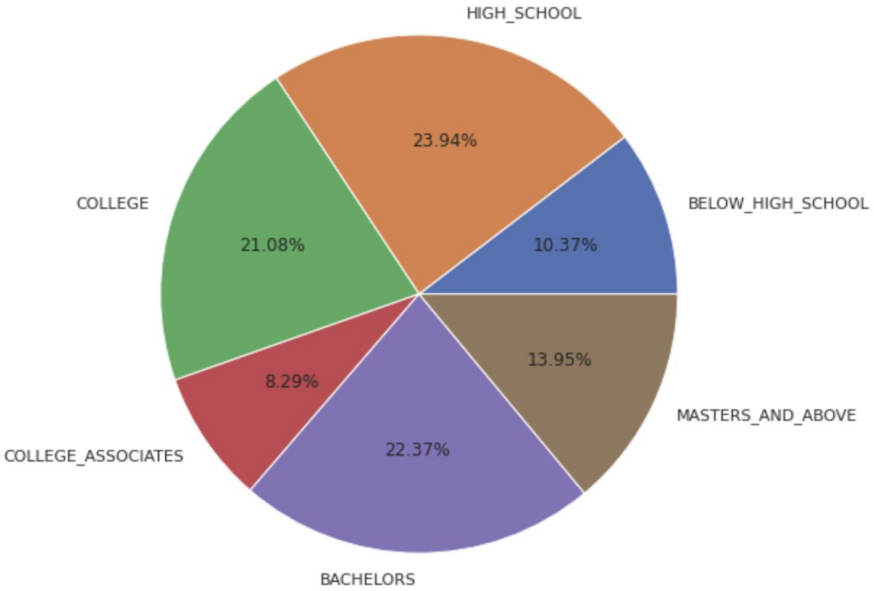
```
Index(['STORE_CODE', 'STO_ID', 'MGT_DIV_DSC_TX', 'SLS_SQR_FTG_QY',
       'STO_TOT_SQR_FTG_QY', 'GAS_STN_FL', 'TAP_POP', 'TAP_POP_5Y_FORECAST',
       'TAP_POP_5YFORECAST_PCT_GROWTH', 'TAP_HH', 'TAP_HH_5Y_FORECAST',
       'TAP_HH_5YFORECAST_PCT_GROWTH', 'TAP_HH_POP', 'TAP_GQ',
       'TAP_PCT_GROUP_QUARTERS', 'TAP_PCT_GQ_INSTITUTION',
       'TAP_PCT_GQ_NON_INSTITUTION', 'TAP_PCT_GQ_NON_INSTITUTION_COL',
       'TAP_PCT_GQ_NON_INSTITUTION_MIL', 'TAP_PCT_GQ_NON_INSTITUTION_OTH',
       'TAP_POP_SEASONAL', 'TAP_PCT_SEASONAL_POP', 'TAP_POP_TRANSIENT',
       'TAP_PCT_TRANSIENT_POP', 'TAP_WP_EMPLOYEES', 'TAP_WP_ESTABLISHMENTS',
       'TAP_PCT_HH_HOH_WHITE', 'TAP_PCT_HH_HOH_BLACK', 'TAP_PCT_HH_HOH_ASIAN',
       'TAP_PCT_HH_HOH_HISPANIC', 'TAP_PCT_HH_HOH_OTHER',
       'TAP_PCT_HH_HOH_UNDER25', 'TAP_PCT_HH_HOH_25_44',
       'TAP_PCT_HH_HOH_45_64', 'TAP_PCT_HH_HOH_65PLUS', 'TAP_PCT_HH_SIZE_1',
       'TAP_PCT_HH_SIZE_2', 'TAP_PCT_HH_SIZE_3', 'TAP_PCT_HH_SIZE_4',
       'TAP_PCT_HH_SIZE_5', 'TAP_PCT_HH_SIZE_6', 'TAP_PCT_HH_SIZE_7PLUS',
       'TAP_PCT_HH_FAMILY_WITH_CHILDRE', 'TAP_PCT_HH_FAMILY_NO_CHILDREN',
       'TAP_PCT_HH_NON_FAMILY', 'TAP_PCT_HH_ONE_PERSON', 'TAP_AVG_HH_INCOME',
       'TAP_AVG_HH_DISPOSABLE_INCOME', 'TAP_AVG_HH_DISCRETIONARY_INCOM',
       'TAP_AVG_HH_WEALTH', 'TAP_PCT_HH_INCOME_UNDER15',
       'TAP_PCT_HH_INCOME_15_30', 'TAP_PCT_HH_INCOME_30_40',
       'TAP_PCT_HH_INCOME_40_50', 'TAP_PCT_HH_INCOME_50_70',
       'TAP_PCT_HH_INCOME_70_100', 'TAP_PCT_HH_INCOME_100_150',
       'TAP_PCT_HH_INCOME_150_200', 'TAP_PCT_HH_INCOME_200PLUS',
       'TAP_PCT_WLTH_0_AND_UNDER', 'TAP_PCT_WLTH_1_5', 'TAP_PCT_WLTH_5_10',
       'TAP_PCT_WLTH_10_25', 'TAP_PCT_WLTH_25_50', 'TAP_PCT_WLTH_50_100',
       'TAP_PCT_WLTH_100_250', 'TAP_PCT_WLTH_250_500', 'TAP_PCT_WLTH_OVER_500',
       'TAP_PCT_HH_NO_VEHICLE', 'TAP_PCT_HH_1_VEHICLE', 'TAP_PCT_HH_2_VEHICLE',
       'TAP_PCT_HH_3_VEHICLE', 'TAP_PCT_HH_4_VEHICLE',
       'TAP_PCT_HH_5PLUS_VEHICLE', 'TAP_PCT_POV_HH',
       'TAP_PCT_POV_HH_ABOVE_POV', 'TAP_PCT_EDU_BELOW_HIGH_SCHOOL',
       'TAP_PCT_EDU_HIGH_SCHOOL', 'TAP_PCT_EDU_SOME_COLLEGE',
       'TAP_PCT_EDU_COLLEGE_ASSOCIATES', 'TAP_PCT_EDU_COLLEGE_BACHELORS',
       'TAP_PCT_EDU_MASTERS_AND_ABOVE', 'TAP_PCT_SSA_POP',
       'TAP_PCT_POP_AGE_18UNDER', 'TAP_PCT_POP_AGE_18_25',
       'TAP_PCT_POP_AGE_25_35', 'TAP_PCT_POP_AGE_35_45',
       'TAP_PCT_POP_AGE_45_55', 'TAP_PCT_POP_AGE_55_65',
       'TAP_PCT_POP_AGE_65_75', 'TAP_PCT_POP_AGE_75PLUS',
       'TAP_PCT_POP_AGE_65PLUS', 'TAP_PCT_UNEMPLOYMENT', 'TAP_POP_DENSITY',
       'TAP_URBANICITY'],
      dtype='object')
```

- Race
- Family Size
- Family Income
- Vehicle Number
- Education Level
- Age

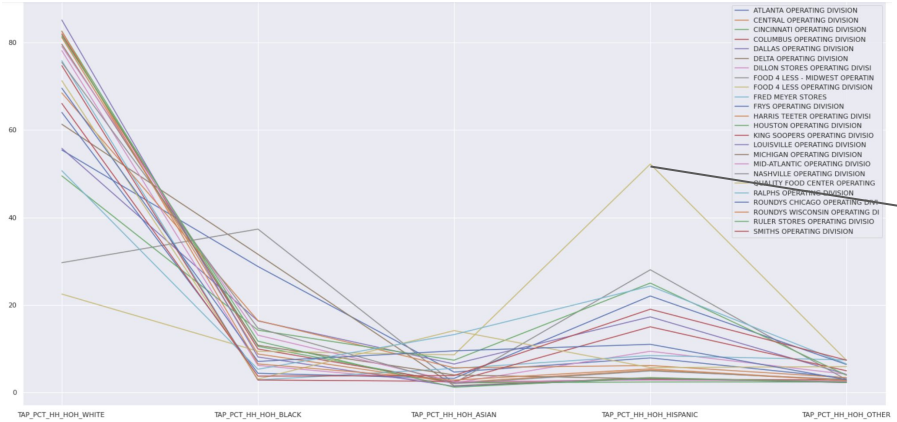
Exploratory Data Analysis - Retailer A

Comparison of percentage distribution of customers





Exploratory Data Analysis - Retailer A

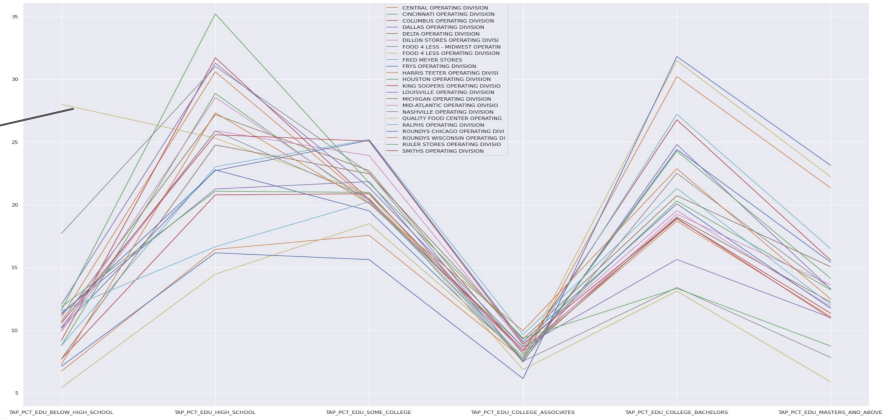


The customer percentage distribution of different races across different groups

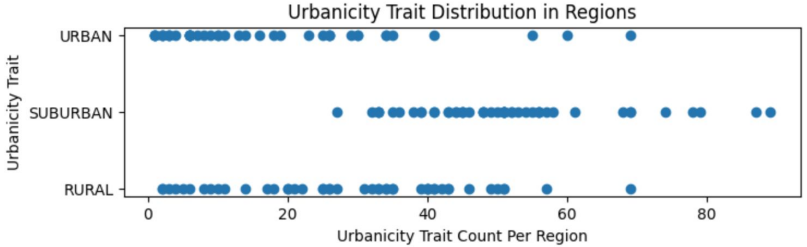
Different behavior

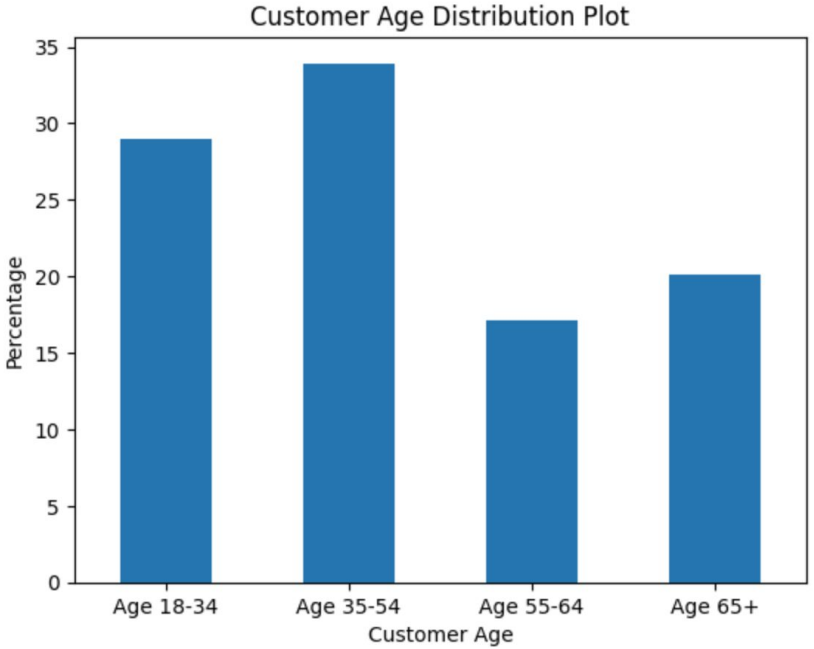
The customer percentage distribution of different education levels across different groups

Different behavior

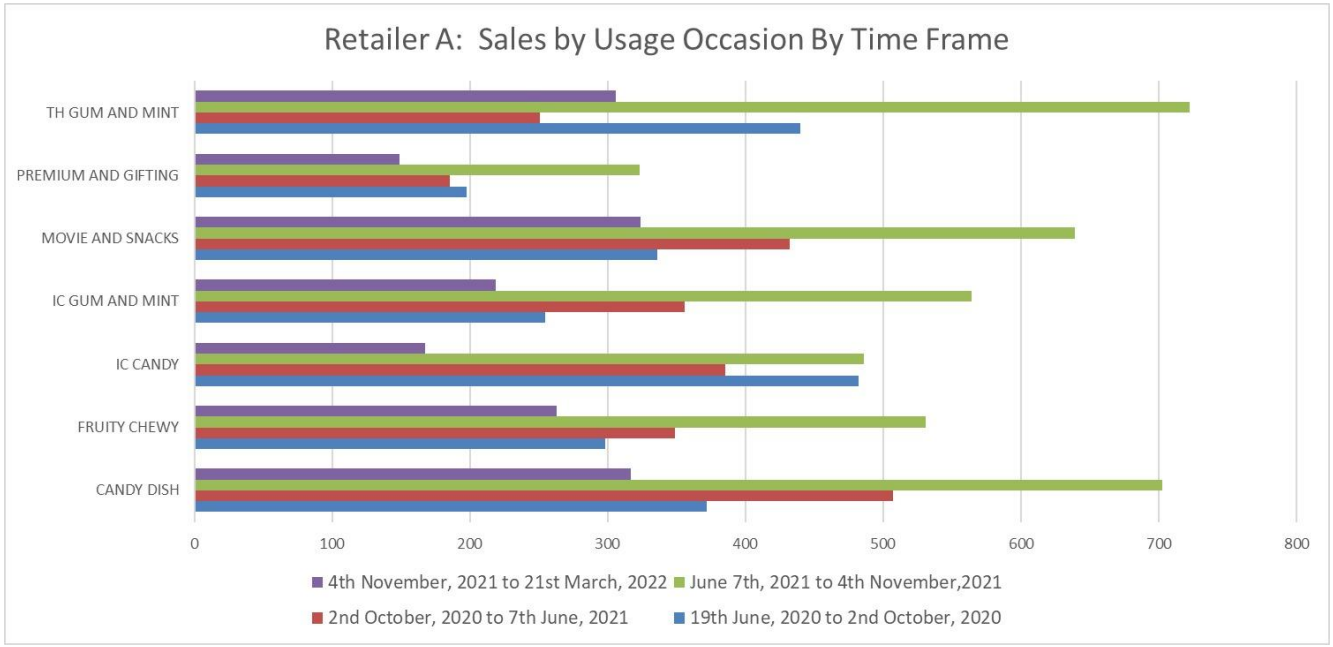


Exploratory Data Analysis - Retailer B





Exploratory Data Analysis



Analyzing **Sales** by **Time frame:**

- Higher sales in **June - Nov 2021** Period for most Usage Occasion (Green bar)

- Lesser sales of **Premium and Gifting** in all timeframes, (less demand, COVID etc.)

- Sales relatively lower during **COVID** for some categories like **Movie & Snacks, Premium & Gifting** (Blue bar and Red bar)

Exploratory Data Analysis

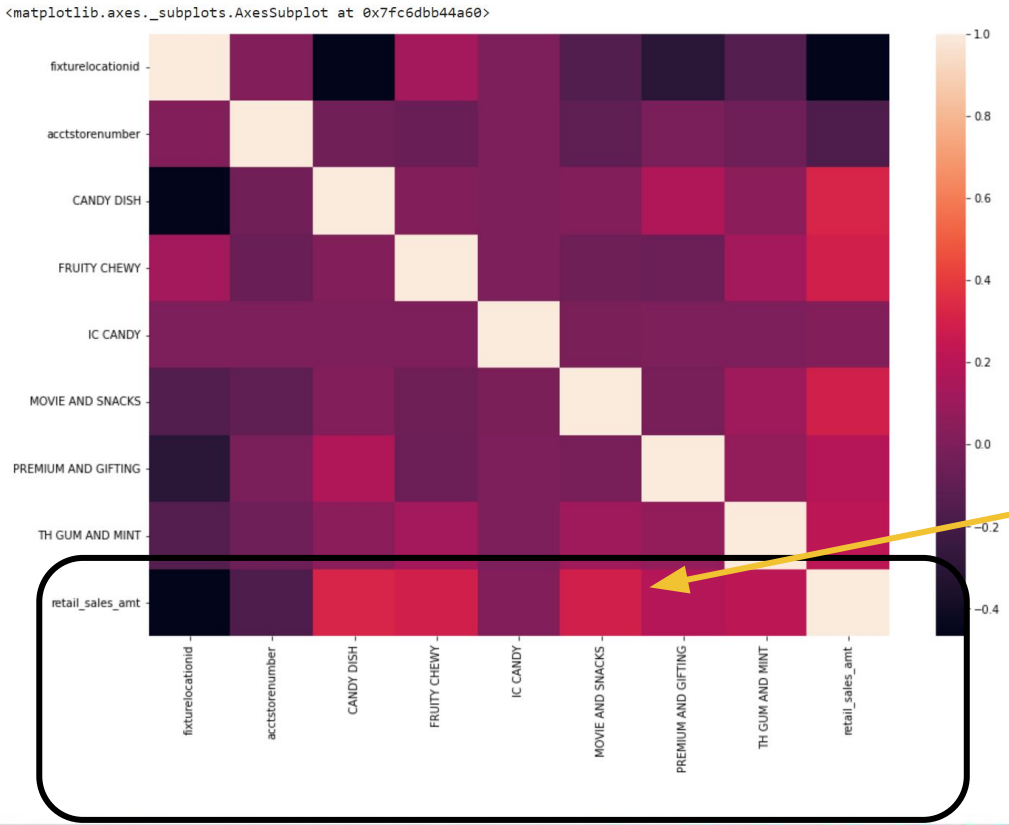
Analyzing **Sales** by **Shelf sizes**

- Should we **optimize** product space on all **shelf sizes**?
- Range:
  144 to 1152 inches
- Custom range:
  1. 100 to 400 -> **SMALL**
  2. 500 to 800 -> **MEDIUM**
  3. 900 to 1200 -> **LARGE**

- **Shelf size ~ Avg Sales**

Shelf Size	Avg Sales
SMALL	**246 M**
MEDIUM	**272 M**
LARGE	**297 M**

Exploratory Data Analysis



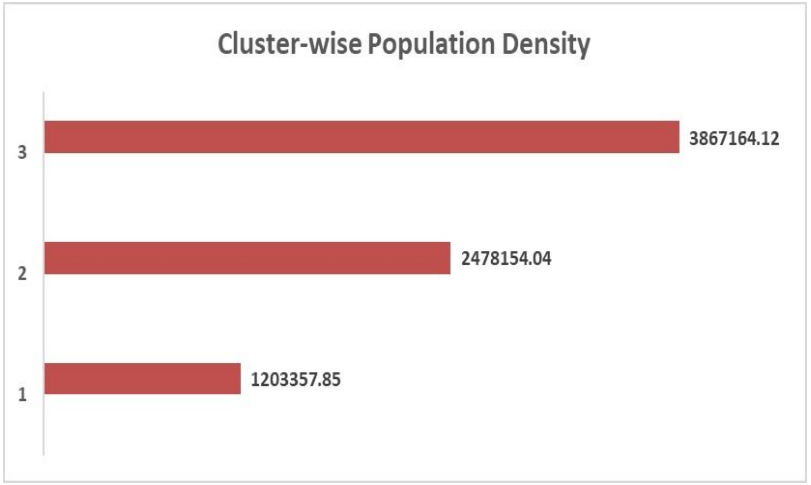
Retail Sales ~ Percentage Spacing
(Red/magenta throughout)

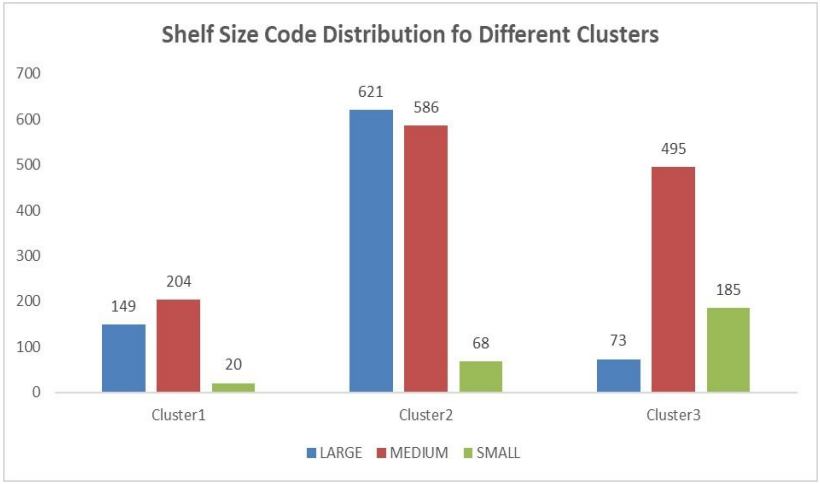
Exploratory Data Analysis

Analyzing **Sales** by **Stores**

- Generalizing **recommendation** for **stores** to increase sales
- Use **Demographic Data** to get more insights about the store similarity
- **K Prototype Clustering** model for further analysis

Exploratory Data Analysis





Exploratory Data Analysis

Current **product placement by usage occasion across Retailers**

- What is the current spacing for each **usage occasion**?

Premium & Gifting	Candy Dish	Movie & Snacks	Take Home Gum & Mint	Fruity Chewy
6%	64%	22%	3%	5%

Time frame1:

13th December, 2021 to 6th April, 2022

Pipeline

Initial Preprocessing

Exploratory Data Analysis

Data Transformation

Master Dataframe

Modeling

Regression to Predict Sales for each shelf

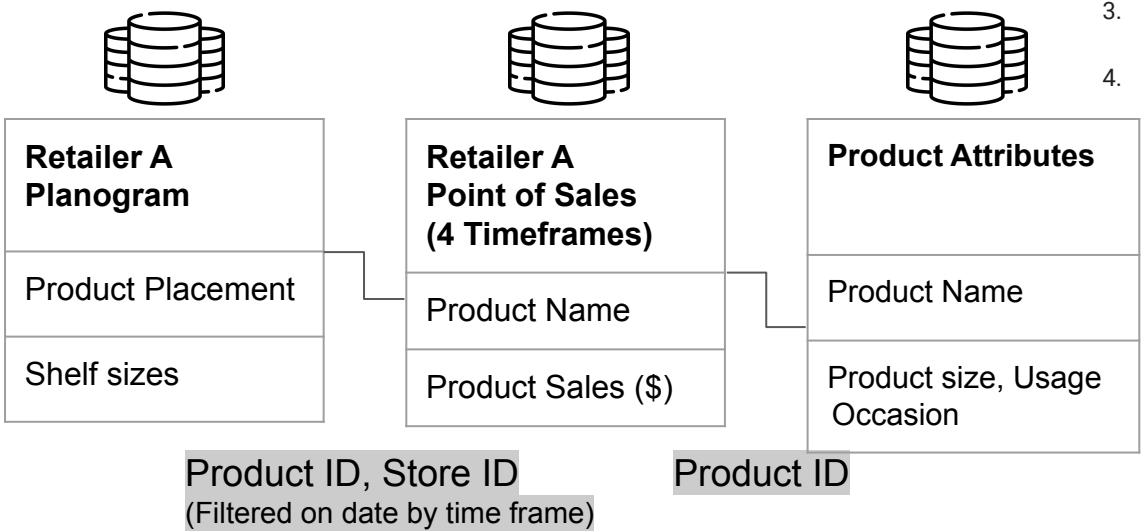
Brute Force Optimization

Shelf % for each usage occasion

Data Transformation

- Checked for NULL values -> Clean data, very few missing values
- Creating a **master data frame**

1. 9th November, 2020 to 12th April, 2021
2. 5th April, 2021 to 17th November, 2021
3. 13th September, 2021 to 5th January,2022
4. 13th December, 2021 to 6th April, 2022



Final Data Frame

- Used Minimum and Maximum of positions to get **Spacings** for each:
  1. Store
  2. Shelf level
  3. Usage Occasion

- Master DataFrame: **Usage Occasion Spacing %**

	fixturelocationid	acctstorenumber	CANDY DISH	FRUITY CHEWY	IC CANDY	MOVIE AND SNACKS	PREMIUM AND GIFTING	TH GUM AND MINT	retail_sales_amt
0	1	1	0.828417	0.0	0.0	0.000000	0.033222	0.018819	8297.099734
1	1	2	0.822764	0.0	0.0	0.217306	0.008500	0.017389	11231.189693
2	1	3	0.308846	0.0	0.0	0.241154	0.000000	0.015016	6208.549898
3	1	4	0.209660	0.0	0.0	0.277007	0.000000	0.013728	6963.789829
4	1	5	0.296080	0.0	0.0	0.284773	0.000000	0.011894	4519.719874

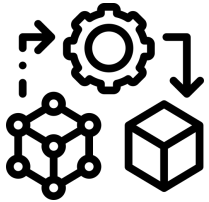
Modeling

1. **Regression**
   - Predicted Sales per shelf by using normalized spacing percentages (features for each usage occasion)

Linear Regression	Random Forest Regressor
- KFold Cross Validation, Ridge Regularization	- Grid Search Cross Validation for hyperparameter tuning

	Linear Regression	Random Forest Regressor
Root Mean Squared Percentage Error	1.35	**0.57**

Optimization



2. Brute Force Optimization

Linear Optimization using Random Forest Regressor

Decision Variables: % spacing allotted to products of each usage occasion for a shelf type:

w1, w2, w3, w4, w5 -> For each usage occasion

Objective: Maximize sales per shelf using Random Forest Model

Constraints:

1. Minimum spacing of 5% usage occasion should be represented on each shelf
2. Shelf should be 100 % full

Model Output

- **Optimal product spacing** (%) for each **usage occasion and for each shelf type**
- Expected Sales**: $9000 per shelf** in 5-6 months.

Premium & Gifting	Candy Dish	Movie & Snacks	Take Home Gum & Mint	Fruity Chewy
5%	25%	35%	5%	30%

Business Impact

Assuming stores use the optimized spacing we recommended:

- Overall total sales before optimization:**0.23 billion USD**
- After Optimization: **0.27 billion** USD, a **17% increase**!

More granular optimization: per shelf type (SMALL, MEDIUM, LARGE):
Small: 24% increase, Medium: 12%, Large: 20%

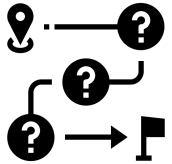
Optimized spacing

1. Increased Visibility

2. Increased Brand Loyalty

3. More $$

Challenges



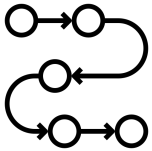
- **Processing:** Certain types of joins lead to loss of data or data redundancy.

- **Generalizability of the solution**: Can the same type of optimization be applied to all stores? We need to choose the correct level of granularity.

- **Temporal Variation:** Trying to determine how our solution would vary across different time periods. Need different optimization methods would need to be used for different time frames.

- **Interpretability/Communication**: Communicating results to the retailer in an effective way

Next Steps



1. Using **store-wise EDA/clustering** to identify groups of stores with similar features, run optimization for those groups separately.

2. **Product family(~top 5 products) wise spacing** (%) for each **usage occasion**

3. Conduct **AB Testing** for different store subgroups with very similar features to establish a **causal relationship** between space optimization and sales.

Sources

1. **Slide 3:** https://www.cspdailynews.com/snacks-candy/hershey-shines-candy-mint-gum-snack-sales
2. **Slide 14:** https://medium.com/@AlainChabrier/shelf-space-optimization-280903f8d1f5
3. **Slide 14:** https://www.analyticsvidhya.com/blog/2016/09/a-beginners-guide-to-shelf-space-optimization-using-linear-programming/
4. **Slide 15:** https://www.analyticsvidhya.com/blog/2017/02/lintroductory-guide-on-linear-programming-explained-in-simple-english/
5. **Slide 15:** Maximizing Profit via Assortment and Shelf-Space Optimization for Two-Dimensional Shelves, Alexander Hübner,Fabian Schäfer,Kai N. Schaal, 22 September 2019, https://onlinelibrary.wiley.com/doi/full/10.1111/poms.13111

Thank you!
Questions?